August, 1965

VARIANCE BOUNDS AND ORTHOGONAL EXPANSIONS IN HILBERT SPACE WITH AN APPLICATION TO INEQUALITIES FOR GAMMA FUNCTIONS AND π

Harold Ruben

Technical Report No. 61

University of Minnesota
Minneapolis, Minnesota

VARIANCE BOUNDS AND ORTHOGONAL EXPANSIONS IN HILBERT SPACE WITH AN APPLICATION TO INEQUALITIES FOR GAMMA FUNCTIONS AND π*

HAROLD RUBEN, University of Minnesota

1. Introduction and summary.

Wallis' inequalities

$$(1) \qquad \frac{2}{2\ell+1}\alpha_\ell < \pi < \frac{1}{\ell}\alpha_\ell \qquad (\ell = 1,2,\ldots),$$

$$\alpha_\ell = \frac{2^2 \cdot 4^2 \cdot 6^2 \cdot \ldots \cdot (2\ell)^2}{1^2 \cdot 3^2 \cdot 5^2 \cdot \ldots \cdot (2\ell-1)^2}$$

have recently been sharpened to

$$(2) \qquad \frac{4\ell+3}{(2\ell+1)^2}\alpha_\ell < \pi < \frac{4}{4\ell+1}\alpha_\ell$$

by Gurland [5], who also obtained the more general result†

$$(3) \qquad \frac{\Gamma(c-2b)\Gamma(c)}{\Gamma^2(c-b)} \geq \frac{b^2+c}{c} \qquad (b,c \text{ real};\ c > 0,\ c-2b > 0),$$

with equality holding for $b=0,-1$. (Gurland as well as subsequent authors (Olkin [9], Erber [3]), give $c > 0$, $c-b > 0$ as the conditions, but this is incorrect if $b > 0$, while for $b \leq 0$ the condition $c > 0$ renders the further condition $c-b > 0$ empty.) The upper and lower bounds in (2) follow from (3) on setting $b = -\frac{1}{2}$, $c = \ell$ and $b = -\frac{1}{2}$, $c = \ell+\frac{1}{2}$.

Although (3) is purely analytic in character, the derivation was based on a fundamental statistical property, namely the Cramér-Rao lower bound for the variance of a regular unbiased estimator of a parameter, here $1/\theta^b$, where θ denotes the scale parameter in a gamma distribution with specified shape parameter c. Similarly (although this has not in fact been carried through in the previously quoted references), successively sharper results can be obtained with

---

*Research supported under a grant from the United States Air Force Office of Scientific Research - Office of Aerospace Research, Grant #AF-AFOSR-885-65.

†Gurland's α and λ correspond in the present notation to -b and c, respectively.

the aid of the more stringent, Bhattacharyya, lower bounds developed in the general theory of statistical estimation. However, the derivation of (3) and its extensions (see Section 3) within a statistical framework is artificial and obscures the significance of the results. Actually, the latter results are essentially Bessel's inequality applied to mean-square approximation of $1/x^b$ $(x > 0)$ in terms of $\{L_i^{(c-1)}(x)\}$, the set of Laguerre polynomials, as a basis. This more natural approach identifies the partial sums of a hypergeometric series as successive lower bounds for the left member of (3) and, in particular, yields a new constellation of infinite series for π and $1/\pi$, with concomitant lower and upper bounds for π. In a more general setting, it is hoped that the discussion (Section 2) of the tie-up between variance bounds for estimators and the expansion of functions (in L_2) in terms of orthogonal functions may be of interest to mathematicians with no statistical orientation.

Notation. (i) $(x)_y$ shall denote $\Gamma(x+y)/\Gamma(x)$.

(ii) i,j shall assume, unless otherwise specified, all non-negative integral values, and k,ℓ all positive integral values.

2. Variance bounds and orthogonal expansions.

The Cramér-Rao and Bhattacharyya counds (Cramér [2], Rao [10], Bhattacharyya [1], Lehmann [8]) constitute an infinite set of non-decreasing lower bounds (of which the Cramér-Rao is the first) for the variance of an unbiased estimator under suitable regularity conditions. These bounds are intimately related to classical Hilbert space analysis, more specifically to Bessel's inequality, with the one additional (fundamentally non-restrictive but practically complicating) feature that the coordinate dunctions $\varphi_0(\equiv 1)$, $\varphi_1, \ldots,$ assumed (as always) to be linearly independent, are seldom orthogonal with respect to the weight function under consideration. (Orthogonalization can, of course, be accomplished by, for example, the Gram-Schmidt process, but this is seldom attempted in the statistical context.) For arbitrary (not necessarily orthogonal) φ_i, the best approximation, in the usual mean-square sense, of a given square-integreble function f by the linear combination $\sum_{0}^{k-1} a_i\varphi_i$ is achieved (using the customary inner product notation) for[†]

$$(4) \qquad a_i^*(k) = \sum_{j=0}^{k-1} w^{ij}(k)(f,\varphi_j) \qquad (i = 0,1,\ldots,k-1) \quad ,$$

where

$$w^{ij}(k) = (w_{ij})^{-1} \quad , \quad w_{ij} = (\varphi_i,\varphi_j) \qquad (i,j = 0,1,\ldots,k-1) \quad .$$

In fact, from the orthogonality relation

$$(f - \sum_{0}^{k-1} a_i^*(k)\varphi_i,\ \varphi_j) = 0 \qquad (j = 0,1,\ldots,k-1) \quad ,$$

the decomposition

$$f - \sum_{0}^{k-1} a_i\varphi_i = (f - \sum_{0}^{k-1} a_i^*(k)\varphi_i) + \sum_{0}^{k-1}(a_i^* - a_i)\varphi_i \quad ,$$

with the a_i arbitrary, leads to

$$\| f - \sum_{0}^{k-1} a_i \varphi_i \|^2 = \|f - \sum_{0}^{k-1} a_i^*(k)\varphi_i \|^2 + \sum_{i,j=0}^{k-1} (\varphi_i,\varphi_j)(a_i - a_i^*(k))(a_j - a_j^*(k))$$

($\| g\|^2 = (g,g)$) , the minimizing a_i being thus $a_i^*(k)$. (The second term on the right, being a positive definite function of the $a_i - a_i^*(k)$, is minimized for $a_i - a_i^*(k) = 0$.) On setting $a_i = a_i^*$, we obtain

$$(5) \qquad \| f\|^2 = \| f - \sum_{0}^{k-1} a_i^*(k)\varphi_i\|^2 + \sum_{i,j=0}^{k-1} (\varphi_i,\varphi_j)a_i^*(k)a_j^*(k) \quad ,$$

---

† The notation $w^{ij}(k)$, $a_i^*(k)$ highlights the dependence of $\{w^{ij}\}$ (though not of $\{w_{ij}\}$) and of $\{a_i^*\}$ on k in the case of non-orthogonal $\{\varphi_i\}$.

whence

$$\|f\|^2 \geq \sum_{i,j=0}^{k-1} (\varphi_i,\varphi_j)\; a_i^*(k) a_j^*(k)$$

(6)

$$= \sum_{i,j=0}^{k-1} w^{ij}(k)(f,\varphi_i)(f,\varphi_j) = B_{k-1} \quad \text{(say)} ,$$

with equality attained if, and only if, $f = \sum_0^{k-1} a_i^*(k)\varphi_i$ almost everywhere with respect to the given weight function. Again, from the minimizing property of the $a_i^*(k)$,

$$\| f - \sum_0^{k-1} a_i^*(k+1)\varphi_i \|^2 \leq \| f - \sum_0^{k-1} a_i^*(k)\varphi_i \|^2 ,$$

whence, from (5) and (6),

(7) $$B_0 \leq B_1 \leq B_2 \leq \ldots .$$

Formulae (4) and (6) are slight extensions of familiar results, in the sense that for an orthonormal system of φ_i, (4) reduces to

(4') $$a_i^* = (f,\varphi_i) ,$$

the Fourier coefficients of f with respect to $\{\varphi_i\}$, and (6) reduces to Bessel's inequality

(6') $$\| f\|^2 \geq \sum_0^{k-1} a_i^{*2} .$$

For a complete orthonormal system of φ_i, $\geq$ in (6) (as in (6')) can be replaced by $=$ and $k-1$ by ∞, to give the corresponding extension of Parseval's formula.

We remark that the Bessel bounds B_{k-1} (being the difference between $\|f\|^2$ and $\min_{a_i} \|f - \sum_0^{k-1} a_i\varphi_i\|^2$) are clearly invariant under non-singular linear transformations on the φ_i of the form $\varphi_i' = \sum_{j=0}^{k-1} c_{ij}\varphi_j$ $(i=0,1,\ldots,k-1;\ k=1,2,\ldots)$.

[The invariantive property is also easily proved directly.] In particular, the B_{k-1} can be computed from (6') by choosing an orthonormal $\{\varphi_i\}$.

Turning now to the statistical aspects, the Cramér-Rao and Bhattacharyya bounds constitute an application of (6) in which

$$\varphi_i = \frac{1}{p_\theta(x)} \frac{\partial^i}{\partial\theta^i} \; p_\theta(x) \;,$$

with θ an unknown parameter in an open interval I on the real line, and $p_\theta(\cdot)$ the density of a (real or vector-valued) random variable X with respect to a fixed (θ-free) σ-finite measure μ defined on an additive class of subsets in the domain of X (in statistical applications, the domain can typically be identified, with no essential loss in generality, as a finite-dimensional Euclidean space, and μ is either Lebesgue or counting measure), while $f(X)$ is a real-valued, unbiased, finite second-moment estimator of a specified parametric function $\tau(\theta)$. Thus, for $\theta \in I$,

$$(8) \qquad \int p_\theta(x)d\mu(x) = 1, \quad (f,\varphi_0) \equiv \int f(x)p_\theta(x)d\mu(x) = \tau(\theta).$$

Under the assumptions that, for each $\theta \in I$, (i) the φ_i are linearly independent (ii) $\partial^i p_\theta(x)/\partial\theta^i$ exists for almost all $(\mu)x$ (iii) $\tau(\theta)$ has finite derivatives of all order (iv) k-fold differentiation under the integral signs in (8) is permissible for $k = 1,2,\ldots,$ (8) gives

$$(\varphi_0,\varphi_i) = \delta_{i0} \;, \quad (f,\varphi_i) = \frac{d^i}{d\theta^i}\,\tau(\theta) \;,$$

so that the lower bounds for $||f||^2$, the second moment of f, are

$$B_{k-1} = \sum_{i,j=0}^{k-1} w^{ij}(k) \frac{d^i}{d\theta^i} \tau(\theta) \quad \frac{d^j}{d\theta^j} \tau(\theta)$$

(9)

$$= \tau^2(\theta) + \sum_{i,j=1}^{k-1} w^{ij}(k) \frac{d^i}{d\theta^i} \tau(\theta) \quad \frac{d^j}{d\theta^j} \tau(\theta)$$

$(B_0 = \tau^2(\theta))$, and, correspondingly, the lower bounds† for $\| f - (f,\varphi_o) \|^2$, the variance of f, are

(10) $$0,\ B_1 - B_0,\ B_2 - B_0,\ \ldots \quad .$$

$B_1 - B_0 = w''(2)(d\tau(\theta)/d\theta)^2$ is the Cramér-Rao bound [$w''(2) = 1/w_{11}$ is the reciprocal of the celebrated Fisherian information function (Fisher, [4], Kendall and Stuart [7]), defined by $E(\partial \log p_\theta(x)/\partial\theta)^2]$, and for $k > 2$ $B_{k-1} - B_0$ is the (k-2)th Bhattacharyya bound for the variance of f. Note that for $\tau(\theta) = \theta$,

(11) $$B_{k-1} - B_0 = w''(k).$$

3. Inequalities for gamma functions and π. In Section 2, specialize μ, p_θ, I and f as follows: μ is Lebesgue measure on the line, I is $(0, \infty)$,

$$p_\theta(x) = \begin{cases} \dfrac{1}{\Gamma(c)\theta^c} e^{-\frac{x}{\theta}} x^{c-1} & (x > 0), \\ 0 & (x \leqq 0), \end{cases} \qquad (c > 0)$$

† The discussion shows that the bounds for the second moment and variance of f are merely statements, couched in statistical terminology, of the fact (consequent on the non-negativeness of Gram matrices) that the determinant of a Gram matrix is necessarily $\geqq 0$. $B_{k-1} - B_0$ corresponds to a Gram matrix of order k, consisting of (w_{ij}), $i,j = 1,\ldots,k-1$, and a bordered row and column with elements $(f-(f,\varphi_0), f-(f,\varphi_0))$, $(f,\varphi_1),\ldots,(f,\varphi_{k-1})$, with k = 2 corresponding, in particular, to the Cauchy-Schwarz inequality, while B_{k-1} corresponds to a Gram determinant of order k+1, consisting of (w_{ij}), $i,j = 0,1,\ldots,k-1$, and a bordered row and column with elements (f,f), $(f,\varphi_0),\ldots,(f,\varphi_{k-1})$.

$$f(X) = \frac{1}{(c)_{-b}\, X^b} \qquad (X > 0;\ b < c/2),$$

with f defined arbitrarily for $X \leqq 0$. Here the condition $b < c/2$ ensures $\| f \|^2 < \infty$. We find

$$\tau(\theta) = \frac{1}{\theta^b},$$

$$\| f \|^2 = \frac{\Gamma(c-2b)\Gamma(c)}{\Gamma^2(c-b)} \frac{1}{\theta^{2b}},$$

$$\varphi_i = \frac{(-1)^i}{\theta^i} \{i!(c)_i\}^{\frac{1}{2}} \tilde{L}_i^{(c-1)} (\tfrac{x}{\theta}),$$

where $\{\tilde{L}_i^{(c-1)}(x/\theta)\}$ is the system of orthonormal Laguerre polynomials relative to $p_\theta(x)$, i.e.,

$$\tilde{L}_i^{(c-1)}(\tfrac{x}{\theta}) = \left\{\frac{i!}{(c)_i}\right\}^{\frac{1}{2}} L_i^{(c-1)}(\tfrac{x}{\theta}).$$

Here φ_i has been computed from the generating function formula

$$(1-t)^{-c}\, e^{-\frac{xt}{1-t}} = \sum_0^\infty L_i^{(c-1)}(x)\, t^i,$$

which becomes, on replacing x by x/θ and t by $-t/\theta$,

$$(\theta+t)^{-c}\, e^{-\frac{x}{\theta+t}} = \sum_0^\infty \left\{ \frac{(-1)^i}{\theta^{i+c}}\, e^{-\frac{x}{\theta}} L_i^{(c-1)} (\tfrac{x}{\theta})\right\} t^i,$$

or

$$\frac{\partial^i}{\partial t^i}\left\{(\theta+t)^{-c}\, e^{-\frac{x}{\theta+t}}\right\}\Big|_{t=0} \equiv \frac{\partial^i}{\partial \theta^i} (\theta^{-c}\, e^{-\frac{x}{\theta}})$$

$$= \frac{(-1)^i i!}{\theta^{i+c}}\, e^{-\frac{x}{\theta}} L_i^{(c-1)} (\tfrac{x}{\theta}).$$

The last formula is equivalent to

$$\frac{1}{p_\theta(x)} \frac{\partial^i p_\theta(x)}{\partial\theta^i} = \frac{(-1)^i i!}{\theta^i} L_i^{(c-1)} (\tfrac{x}{\theta}) ,$$

which is the previously stated formula for φ_i. Thus

$$\frac{d^i}{d\theta^i} \tau(\theta) = \frac{(-1)^i (b)_i}{\theta^{b+i}} , \quad w_{ij} = \frac{i!(c)_i}{\theta^{2i}} \delta_{ij}$$

and (9) [or (6), with $a_i^* = (1/w_{ii}) \cdot d^i\tau(\theta)/d\theta^i$] gives

$$B_{k-1} = \left(\sum_{i=0}^{k-1} \frac{1}{i!} \frac{\{(b)_i\}^2}{(c)_i} \right) \frac{1}{\theta^{2b}} ;$$

that is,

$$\frac{\Gamma(c-2b)\Gamma(c)}{\Gamma^2(c-b)} \geq \sum_{i=0}^{k-1} \frac{1}{i!} \frac{\{(b)_i\}^2}{(c)_i} \quad (b,c \text{ real}; \ c > 0, \ c-2b > 0), \tag{12}$$

with equality holding if, and only if, $b = 0, -1, \ldots, -(k-1)$. We thus have a sequence of rational function bounds for the function $\Gamma(c-2b)\Gamma(c)/\Gamma^2(c-b)$. Gurland's result†

$$\frac{\Gamma(c-2b)\Gamma(c)}{\Gamma^2(c-b)} \geq \frac{c+b^2}{c}$$

is the case $k = 2$. The significance of this last result and its extensions in (12)

---

† Gurland reparametrizes by choosing $\theta' = 1/\theta^b$ as the parameter (in which case $\tau = \theta'$), rather than θ itself. However, such a reparametrization is inappropriate from the point of view of computing the B_{k-1} (Garland computes B_1 only, using (11)), since the resulting $\varphi_i(\cdot)$, which are linearly related, with coefficients depending on θ only, to the original $\varphi_i(\cdot)$, are non-orthogonal with respect to $q_{\theta'}(\cdot) \equiv p_\theta(\cdot)$, and the computation of the successive B_{k-1} (using (11) again) becomes prohibitively tedious. (The B_{k-1} themselves are unaffected, as pointed out after (6').)

reflect the Bessel lower bounds for $||x^{-b}||^2$, corresponding to mean-square approximation of x^{-b} $(x > 0)$ by successively larger number of Laguerre polynomials $L_i^{(c-1)}(x)$. In particular, $k = 1$ corresponds to approximation by a constant function, and $k = 2$ (Gurland's case) to approximation by a linear function.

From the completeness of $\{L_i^{(c-1)}(\cdot)\}$, we have also

$$(13) \qquad \frac{\Gamma(c-2b)\Gamma(c)}{\Gamma^2(c-b)} = \sum_0^\infty \frac{1}{i!} \frac{\{(b)_i\}^2}{(c)_i} \qquad (b,c \text{ real; } c > 0,\ c-2b > 0).$$

The series in (13) is just the series for the hypergeometric function with unit argument, numerator parameters b, b, and denominator parameter c. This suggests a very simple, direct proof of (12). In fact, on setting $a = b$ in Gauss' formula

$$F(a,b;c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \qquad (R(c-a-b) > 0;\ c \neq 0, -1, \ldots),$$

we obtain

$$(14) \qquad \frac{\Gamma(c-2b)\Gamma(c)}{\Gamma^2(c-b)} = \sum_0^\infty \frac{1}{i!} \frac{\{(b)_i\}^2}{(c)_i} \qquad (R(c-2b) > 0;\ c \neq 0, -1, \ldots).$$

If, further, b, c are real with $c > 0$, then the individual terms of the series in (14) are non-negative, from which (12) follows.

We record here some noteworthy series, with their consequential bounds, for π and $1/\pi$ which derive from (12) and (13). For $b = -\frac{1}{2}$, $c = \ell$ and $b = -\frac{1}{2}$, $c = \ell + \frac{1}{2}$, the left member of (13) reduces to $(\alpha_\ell/\ell)/\pi$ and $\{(\ell + \frac{1}{2})/\alpha_\ell\}\pi$, respectively. ($\alpha_\ell$ is defined after (1).) Accordingly,

$$(15) \qquad \begin{aligned} \frac{1}{\pi} &= \frac{\ell}{\alpha_\ell} \sum_0^\infty \frac{1}{i!} \frac{\{(-\frac{1}{2})_i\}^2}{(\ell)_i} \\ &= \frac{\ell}{\alpha_\ell} \left(1 + \frac{1}{4}\frac{1}{\ell} + \frac{1}{32}\frac{1}{\ell(\ell+1)} + \frac{3}{128}\frac{1}{\ell(\ell+1)(\ell+2)} + \frac{75}{2048}\frac{1}{\ell(\ell+1)(\ell+2)(\ell+3)} + \cdots\right), \end{aligned}$$

and

$$\pi = \frac{\alpha_\ell}{\ell+\frac{1}{2}} \sum_0^\infty \frac{1}{i!} \frac{\{(-\frac{1}{2})_i\}^2}{(\ell+\frac{1}{2})_i}$$

(16)

$$= \frac{2\alpha_\ell}{2\ell+1}\left(1+\frac{1}{2}\frac{1}{2\ell+1} + \frac{1}{8}\frac{1}{(2\ell+1)(2\ell+3)} + \frac{3}{16}\frac{1}{(2\ell+1)(2\ell+3)(2\ell+5)} + \frac{75}{128}\frac{1}{(2\ell+1)(2\ell+3)(2\ell+5)(2\ell+7)}+\cdots\right),$$

in which each positive integral ℓ generates an infinite series for π and for $1/\pi$. [(16) is valid also for $\ell = 0$ if α_0 is defined to be 1.] The partial sums in (15) and (16) provide successively sharper lower bounds for $1/\pi$ and π, respectively, i.e., successively sharper upper and lower bounds for π, and each set of bounds converges to the latter quantity as $\ell \to \infty$. The first three sets of lower and upper bounds are

$$\frac{2}{2\ell+1}\alpha_\ell < \pi < \frac{1}{\ell}\alpha_\ell , \tag{17}$$

$$\frac{4\ell+3}{(2\ell+1)^2}\alpha_\ell < \pi < \frac{4}{4\ell+1}\alpha_\ell , \tag{18}$$

$$\frac{32\ell^2+72\ell+37}{4(2\ell+1)^2(2\ell+3)}\alpha_\ell < \pi < \frac{32(\ell+1)}{32\ell^2+40\ell+9}\alpha_\ell , \tag{19}$$

(17) and (18) giving the Wallis and Gurland bounds.

REFERENCES

1. A. Bhattacharyya, On some analogues of the amount of information and their uses in statistical estimation, Sankhyā, Vol. 8 (1946-7-8).

2. H. Cramér, Mathematical Methods of Statistics, Princeton University Press (1946).

3. T. Erber, The gamma function inequalities of Gurland and Gautschi, Skand. Akt., Vol. 43 (1960).

4. R. A. Fisher, Theory of statistical estimation, Proc. Camb. Phil. Soc., Vol. 22 (1925).

5. J. Gurland, On Wallis' formula, Amer. Math. Monthly, Vol. 63, No. 9, Part 1 (1956).

6. J. Gurland, An inequality satisfied by the gamma function, Skand. Akt., Vol. 39 (1956).

7. M. G. Kendall and A. Stuart, The Advanced Theory of Statistics, Vol. 2, Chas. Griffin, London (1961).

8. E. L. Lehmann, Notes on the Theory of Estimation [recorded by C. Blyth], University of California (1949-50).

9. I. Olkin, An inequality satisfied by the gamma function, Skand. Akt., Vol. 41 (1958).

10. C. R. Rao, Information and accuracy attained in the estimation of statistical parameters, Bull. Calcutta Math. Soc., Vol. 37 (1945).